Local low rank approximation with a parametric disparity model for light field compression

Elian Dib, Mikael Le Pendu, Xiaoran Jiang, Christine Guillemot

Local low rank approximation with a parametric disparity model for light field compression

Elian Dib*, Mikael Le Pendu†, Xiaoran Jiang*, Christine Guillemot* *Fellow, IEEE*

*INRIA Rennes Bretagne Atlantique
Campus de Beaulieu
Rennes, Ille-et-Vilaine, 35 042, France
christine.guillemot@inria.fr

†V-SENSE
School of Computer Science and Statistics
Trinity College Dublin, Ireland
lependum@scss.tcd.ie

***Abstract*—We address the problem of light field dimensionality reduction for compression. We describe a local low rank approximation method using a parametric disparity model. The local support of the approximation is defined by super-rays. A super-ray can be seen as a set of super-pixels that are coherent across all light field views. A dedicated super-ray construction method is first described that constrains the super-pixels forming a given super-ray to be all of the same shape and size, dealing with occlusions. This constraint is needed so that the super-rays can be used as supports of angular dimensionality reduction based on low rank matrix approximation. The light field low rank assumption depends on how much the views are correlated, i.e. on how well they can be aligned by disparity compensation. We first introduce a parametric model describing the local variations of disparity within each super-ray. We then consider two methods for estimating the model parameters. The first method simply fits the model on an input disparity map. We then introduce a disparity estimation method using a low rank prior. This method alternatively searches for the best parameters of the disparity model and of the low rank approximation. We assess the proposed disparity parametric model, first assuming that the disparity is constant within a super-ray, and second by considering an affine disparity model. We show that using the proposed disparity parametric model and estimation algorithm gives an alignment of super-pixels across views that favours the low rank approximation compared with using disparity estimated with classical computer vision methods. The low rank matrix approximation is computed on the disparity compensated super-rays using a singular value decomposition (SVD). A coding algorithm is then described for the different components of the proposed disparity-compensated low rank approximation. Experimental results show performance gains, with a rate saving going up to 92.61%, compared with the JPEG Pleno anchor, for real light fields captured by a Lytro Illum camera. The rate saving goes up to 37.72% with synthetic light fields. The approach is also shown to outperform an HEVC-based light field compression scheme.**

I. Introduction

Light field imaging has emerged as a promising technology for a variety of multimedia applications. To give only a few examples, by capturing light rays emitted by the scene according to different orientations, light fields allow reconstructing 3D models of the scene. They offer the possibility of viewing the scene from any viewpoint and direction of gaze, thus enabling immersive experience in virtual reality (VR) applications, using VR headsets or head mounted displays. Light fields also find applications in biometry, e.g. for face recognition, for object detection, classification and recognition, in computational photography, and in medical imaging with for example 3D light field microscopy. However, light fields represent very large volumes of high dimensional data, hence the need for designing efficient compression algorithms.

In this paper, we focus on the problem of compression of dense light fields, as those captured by plenoptic cameras, which represent very large volumes of highly redundant data. While a number of methods have already been published in the literature aiming at adapting standardized solutions (in particular HEVC) to light field data as in [1]–[4], here we focus on the problem of reducing the angular dimensions of light fields with a low rank approximation method. A homography-based low-rank approximation method called HLRA has been shown to give very good light field compression performances in [5]. However, the validity of the light field low rank assumption depends on how much the views are correlated, and the alignment performed by HLRA may not be optimal for reducing the rank, as the method uses a small number of homographies per view.

In this paper, we explore the use of local models for both view alignment and for low rank approximation, with the goal of reducing the rank and increasing the performance of light field compression algorithms. The support of the local approximation is defined by super-rays. The concept of super-ray has been initially introduced in [6] as an extension of super-pixels to address the computational complexity issue in light field image processing tasks. The term *super-pixel*, first coined in [7] can be seen as the clustering of image pixels into a set of perceptually uniform regions. Similarly, a super-ray can be seen as the clustering of rays of the light field within and across views, hence corresponding to the same set of 3D points of the imaged scene. The super-rays are used here to better expose redundancy across the different views compared to a global homography-based alignment as done in [5]. In order to successfully exploit redundancies across views using a low rank approximation, the super-rays must group super-pixels which are consistent across the views while being constrained to be of same shape and size, hence the need for adapted disparity estimation and compensation methods.

In this paper, we first propose a method for segmenting the input light field into super-rays that satisfy the above

This work has been supported by the EU H2020 Research and Innovation Programme under grant agreement No 694122 (ERC advanced grant CLIM).

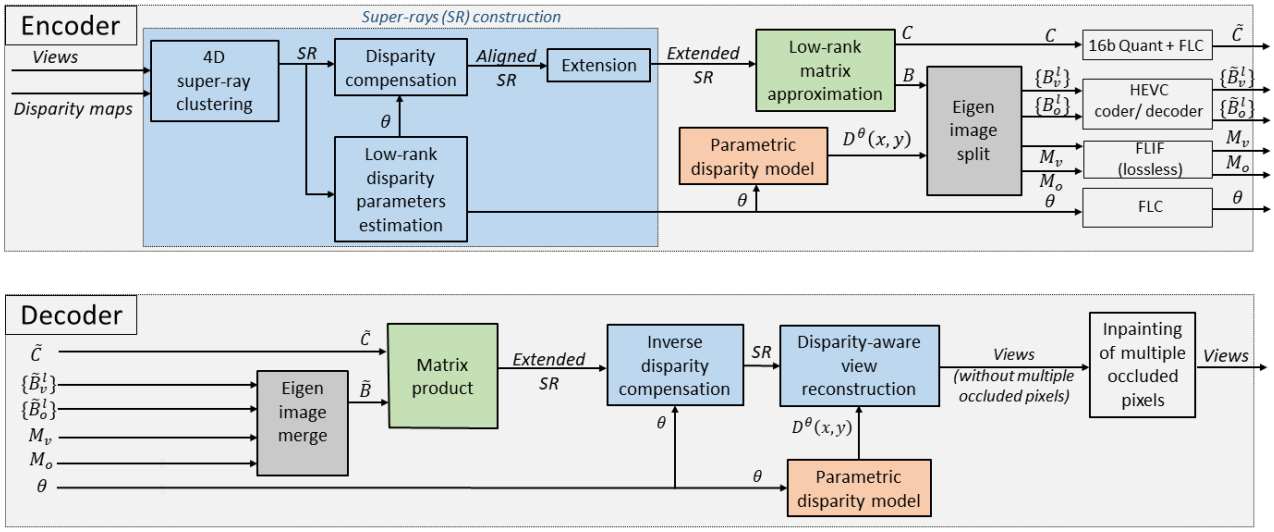

Fig. 1. Overview of the super-ray based low rank approximation, of the coding chain (encoder), and of the light field reconstruction steps (decoder). The eigen images $\mathbf{B}$ are compressed using an HEVC video encoder. The $\mathbf{C}$ and θ coefficients are quantized and then coded using a Fixed Length Code (FLC). Metadata such as the two maps $\mathbf{M}_v$ and $\mathbf{M}_o$ are compressed using the FLIF [8] image encoder.

constraints. The central view is taken as a reference for defining super-ray centroids. The super-rays are formed by minimizing color, disparity and spatial distances between each centroid and the disparity compensated rays. We then address the problem of super-ray based disparity estimation so that the disparity compensation can align the super-rays in a way that will favour the low rank approximation. This work extends the approach presented in [9] by introducing a new local parametric disparity model describing local disparity variations within each super-ray. Two methods are then considered for estimating the model parameters. The first method simply fits the model on an input disparity map. We then formulate the disparity estimation task as a low rank optimization problem: the proposed algorithm iteratively searches for the parameters of the local super-ray disparity model jointly with the low rank matrix of the disparity compensated views. In other words, the proposed disparity estimation method determines the parameters of the disparity model that yield the best low rank approximation, i.e. give the lowest approximation error for a given rank. We assess the proposed disparity parametric model, first assuming that the disparity is constant within a super-ray, and second by considering an affine disparity model. We show that this low rank based disparity estimation method leads to better approximation results with respect to using disparity estimated with a recent computer vision method (in the tests we used [10]). Note that, due to occlusions, individual super-pixels may contain less pixels than the complete area captured in the corresponding super-ray (gathering information from super-pixels of all the views). Therefore, the super-pixels need to be extended with the occluding pixels to satisfy the shape and size constraint.

A low rank approximation is then computed for the set of extended and aligned super-rays that are stacked in a vectorized form in different columns of a matrix $\mathbf{X}$. The rank r approximation of the set of extended super-rays is expressed as a product of a matrix $\mathbf{B}$, containing r columns corresponding to basis (or eigen) images in a vectorized form, with a matrix $\mathbf{C}$ containing weighting coefficients. The matrix $\mathbf{B}$ is re-arranged into two sets of eigen images, respectively corresponding to the light rays visible and occluded in the central view, with their corresponding visible and occluded segmentation maps. We then describe the complete compression algorithm and the tools used for encoding the different components of the approximation model. An image sequence formed with both the sets of 'occluded' and 'visible' eigen images is encoded using HEVC, while the super-ray segmentation maps are encoded with the FLIF coder [8]. The parameters of the disparity model are encoded using simple fixed length codes. Note that the proposed coding scheme extends the method described in [9] by introducing a more general parametric disparity model with the corresponding parameter estimation algorithm guided by a low rank approximation. In addition, further improvements have been brought to super-ray extensions using a low rank matrix completion, and to segmentation map encoding.

The method is compared against two reference schemes: the first scheme encodes all the views as a pseudo video sequence using HEVC-inter coding [11] and the second scheme encodes a low rank approximation of a light field after aligning all the views using homographies [5]. We also compare the PSNR-rate performances of the proposed compression algorithm with the ones obtained with the verification model of JPEG-Pleno (VM 2.0) [12]. Experimental results show that our method compares favorably with the other state of the art methods and that the proposed parameter estimation method for the disparity model is well adapted to our low rank based compression scheme.

II. Related Work

Existing light field compression solutions can be broadly classified into two categories: approaches directly compressing

the lenslet images or approaches coding the views extracted from the raw data.

Methods proposed for compressing the lenslet images mostly extend HEVC intra coding modes by adding new prediction modes to exploit similarity between lenslet images. This is the case in [13], [1] and [2] where the authors use block-based bi-predicted self-similarity techniques, or self-similarity compensated prediction using locally linear embedding (LLE) respectively. A bi-directional mode with vectors coded using advanced motion vector prediction (AMVP) is also introduced in [13]. The authors in [14] further proposed a high-order Intra prediction mode using a geometric transformation applied to a reference block located in the causal region of the block to be predicted. The authors in [15] design a set of predictors for disk-shaped clusters of pixels, called macro-pixels, using an L1 minimization of the prediction residuals. They further propose directional intra-prediction modes based on HEVC for the macro-pixels. Chao et al. [16] directly apply a graph lifting transform on irregularly spaced color components of pixel in the raw data without demosaicing. This avoids redundancy that results from color demosaicing. In [17], a scalable light field coding scheme is proposed, in which the base layer encodes a subset of LF raw data corresponding to a narrow field-of-view (FOV), and the enhancement layers hierarchically encode additional angular information using exemplar-based interlayer prediction. In [18], the plenoptic image is partitioned into tiles, and the sequence of tiles is then compressed using HEVC. The authors of [19] classify the HEVC prediction units (PU) in 3 different categories based on texture homogeneity and use a different prediction mode, based on a Gaussian process regression, for each texture category.

Another category of approaches consists in first extracting sub-aperture images (or views) from the raw plenoptic data, that are then coded as a pseudo-video sequence using HEVC inter coding [11], the latest JEM coder [20], or multi-view HEVC based coding scheme [21] [4]. The authors in [22], propose a coding order and a prediction structure inspired from those used in the multi-view coding (MVC) standard. The author of [23] exploits inter-view correlation by using homography and 2D warping to predict views. Homographies are computed via Random Sample Consensus (RANSAC) [24]. A scalable extension of HEVC-based scheme is also proposed in [25] where a sparse set of micro-lens images (also called elemental images) is encoded in a base layer. The other elemental images are reconstructed at the decoder using disparity-based interpolation and inpainting. The reconstructed images are then used to predict the entire lenslet image and a prediction residue is transmitted yielding a multi-layer scheme.

The authors in [26] and [27] propose hierarchical light field coding structures based on a partitioning of the input sub-aperture images. In [27], a first set of sub-aperture images is compressed as a pseudo video sequence using HEVC, and used to predict the adjacent images by linear interpolation. In [26], the authors use HEVC to encode pseudo sequences of views per quadrant of the light field. They define a hierarchical coding order based on respective view angular positions with a limited number of reference frames adapted to the reference list management of HEVC.

While the above techniques significantly rely on block-based prediction mechanisms of standardized solution, other compression schemes instead use efficient view synthesis techniques for first reconstructing the entire light field from a very sparse set of views [28], [27], [29]. In [28], the authors use a convolutional neural network to predict all views from the four corner ones, while the authors in [29] and [30] rather follow a depth image-based rendering approach in which depth is first estimated and used to warp reference views to predict the others. In [29], the warping is done per segmented region of a reference view (or a set of reference views). The resulting multiple references are then used to predict the others using a sparse predictor. The authors in [30] apply a disparity compensated wavelet coding technique. Disparity-guided sparse coding methods with learned dictionaries are instead considered in [31], while the authors in [27] use a linear approximation computed with Matching Pursuit for disparity based view prediction.

Instead of explicitly using disparity for view prediction, one can also exploit signal priors. This is the case in [32], [33] where the authors exploit light field sparsity in the 4D Fourier domain to reconstruct the entire light field from a subset of views. The approach is assessed using the SHVC (Scalable HEVC-based Video Coding) coding framework with two layers. In [34], light field views are predicted by interpolation using sparsity in the shearlet transform domain. Various models and transforms have also been proposed for light field compression. A global homography-based low rank approximation approach is introduced in [5] while, in [35], the authors describe a framework referred to as Steered Mixture-of-Experts (SMoE) where high-dimensional kernels are used to sparsely represent the plenoptic function. Local transforms applied either on 4D blocks using 4D-DCT [36], or graph-based transforms defined locally on super-rays in [37] have also been explored for light field compression.

In this paper, we explore instead local low rank models on super-rays, and we propose novel parametric disparity estimation methods to favour the low rank approximation.

III. Notations and scheme overview

Let $L(u, v, x, y)$ be a light ray of the light field L, and (u, v, x, y) its coordinates using the two plane parameterisation, where (u, v) and (x, y) are the angular (view) and spatial (pixel) coordinates respectively. A super-pixel $\mathcal{SR}_i$ denotes a group of rays within the same view (u_i, v_i) and a super-ray $\mathcal{SR}$ extends that concept by grouping super-pixels across all views of the light field. In the rest of the paper, $\mathcal{SR} = [\mathrm{vec}(\mathcal{SR}_1) \mid \mathrm{vec}(\mathcal{SR}_2) \mid ...]$ will denote the super-ray matrix formed by vectorizing all super-pixels $\mathcal{SR}_i$ for each view i.

The overall coding scheme is depicted in Fig.1 and comprises the following steps:

- Super-ray construction using a disparity-aided k-means clustering of the rays across all views.
- Low-rank based disparity estimation and compensation of each super-ray, relying on a low rank prior and a disparity parametric model.

- Super-rays extension so that all super-pixels forming a SR are well-aligned and are of same size to allow low rank approximation.
- Low rank approximation of the set of extended super-rays by the product of a matrix $\mathbf{B}$ and a coefficient matrix $\mathbf{C}$.
- Rearranging the matrix $\mathbf{B}$ into a first set $\{\mathbf{B}_v\} = \{\mathbf{B}_v^l\}_{l\in[\![1,r]\!]}$ (with r the approximation rank) corresponding to the light rays visible in the central view, and a second set $\{\mathbf{B}_o\} = \{\mathbf{B}_o^l\}_{l\in[\![1,r]\!]}$, very sparse, and containing eigen values of super-rays extensions.
- Creating two maps $\mathbf{M}_v$ and $\mathbf{M}_o$ to reconstruct respectively the visible and occluded parts of the eigen-super-rays from the eigen-images at the decoder side.
- Encoding the color information to be transmitted to the decoder, i.e., the two sets of eigen images $\{\mathbf{B}_v\}$ and $\{\mathbf{B}_o\}$ and the matrix of coefficients $\mathbf{C}$.
- Encoding the maps $\mathbf{M}_v$ and $\mathbf{M}_o$ to recreate the eigen-super-rays from the sets of eigen-images $\{\mathbf{B}_v\}$ and $\{\mathbf{B}_o\}$ at the decoder side.
- Encoding the parameters $\{\theta\}$ of the disparity model of each super-ray to perform the inverse warping at the decoder side.
- The decoding steps are analogous to the coding steps but processed in the reverse order.

These different steps are detailed below.

IV. Light field over-segmentation

We present a super-ray construction method that groups light rays having similar color and depth, i.e. that correspond to the same set of points in the 3D space. The clustering of the light field into super-rays is performed, as in [6], by computing a distance that combines similarity in terms of colour and disparity as well as the spatial distance between the projected ray on the reference view and the cluster centroid.

Let $r = (x, y, u, v)$ denote the coordinates of a ray (or a pixel) at the spatial coordinates (x, y) in a view of angular coordinates (u, v). Knowing its disparity d, the ray will be projected on the central view at a position $(x_c, y_c) = T^d_{u,v}(x, y)$, where the projection operator $T^d_{u,v}$ is defined as

$$T^d_{u,v} : (x, y) \mapsto (x + d(u_c - u), y + d(v_c - v)). \quad (1)$$

Thus, $r = (x, y, u, v)$ and $r_c = (x_c, y_c, u_c, v_c)$ are imaging the same scene point.

The light field L can then be segmented into clusters. The algorithm proceeds as follows. A set $\mathcal{S}$ of centroids is initialized by taking regularly sampled rays in the central view. Each centroid $s \in \mathcal{S}$ is defined by its spatial coordinates (x_s, y_s) in the central view, its color Lab_s in *CIELab* color space, and its disparity d_s. The color Lab_s and disparity d_s are initialised respectively as the light field color and disparity values at the centroid coordinates (x_s, y_s, u_c, v_c). Note that for the clustering we use an input disparity map that, in the experiment, we computed using the method in [10]. Then the clusters and centroids are alternatively updated until convergence with the following *assignment* and *update* steps:

- *Assignment step:* each ray $r = (x, y, u, v)$ is assigned to the closest centroid s with respect to the distance function $\Delta_{Lab,xy,d}$ defined as

$$\Delta_{Lab,xy,d}(r, s) = \Delta_{Lab} + \lambda_{xy}\Delta_{xy} + \lambda_d\Delta_d, \quad (2)$$
$$\Delta_{Lab}(r, s) = \|Lab_r - Lab_s\|^2, \quad (3)$$
$$\Delta_{xy}(r, s) = \|T^d_{u,v}(x, y) - (x_s, y_s)\|^2, \quad (4)$$
$$\Delta_d(r, s) = \|d_r - d_s\|^2, \quad (5)$$

  where Lab_r and d_r are respectively the color (in *CIELab* colorspace) and the disparity values of the ray r, and where λ_{xy} and λ_d are weights for the spatial and disparity distances. All the rays assigned to the same centroid thus form a cluster called a super-ray.
- *Update step:* for each super-ray, its centroid s is updated by setting Lab_s, d_s and (x_s, y_s) to the averages within the super-ray of respectively, the colors, the disparities, and the spatial coordinates projected on the central view.

The final super-rays thus group rays of the light field that have similar color and depth. This allows us to better align the super-pixels within a super-ray after disparity compensation, which renders low rank approximation more efficient as presented in the next section.

V. Disparity estimation using low rank priors

It was shown in [5] that using homographies to globally align the views reduces the low rank approximation error of light fields with small baselines. However the homographies fail to correctly align the views with large baselines. To address this issue, we propose instead to perform the disparity compensation locally on each super-ray. As using homographies to perform the alignment on each super-ray would significantly increase the amount of side information to be transmitted to the decoder, we propose instead a local parametric model of the disparity variations per super-ray. We describe a disparity estimation method using the low rank approximation error as an optimization criterion of the parameters of the disparity model. More precisely, the proposed method iteratively estimates both the model parameters and the components of the low rank approximation. We show in Section X that estimating the model parameters independently of the low rank constraint, e.g. by fitting the model on the disparity values obtained using [10], is not optimal for our low rank based coding scheme.

A. General problem formulation

Let us first consider the general case where the disparity is a function of the spatial variables (x, y) parameterized by a vector of parameters θ. For the general case, we do not assume the horizontal and vertical disparity to be equal. Hence, we note the horizontal and vertical disparities $D^\theta_x(x, y)$ and $D^\theta_y(x, y)$ respectively. For a given super-ray, the associated parameter vector θ must be transmitted instead of a disparity map (i.e. per-pixel disparity value).

For a given super-ray, let T^θ be the set of disparity compensation operators T_i^θ, where T_i^θ is the operator for the super-pixel in view (u_i, v_i) forming a super-ray. This operator can be defined as

$$T_i^\theta : (x, y) \mapsto \left(X_i^\theta(x, y), Y_i^\theta(x, y)\right), \tag{6}$$

with,

$$X_i^\theta(x, y) = x + D_x^\theta(x, y) \cdot (u_i - u_c), \tag{7}$$

$$Y_i^\theta(x, y) = y + D_y^\theta(x, y) \cdot (v_i - v_c), \tag{8}$$

where c is the index of the central view. This operator matches the pixel (x, y) in the central view (u_c, v_c) to the pixel $T_i^\theta(x, y)$ in view (u_i, v_i). The disparity-compensated super-ray is obtained by applying the operator T_i^θ for each super-pixel $\mathcal{SR}_i$ of the super-ray $\mathcal{SR}$

$$\mathcal{SR}_{i,warped} = \mathcal{SR}_i \circ T_i^\theta. \tag{9}$$

Let $\mathcal{SR} \circ T^\theta = \left[\text{vec}(\mathcal{SR}_1 \circ T_1^\theta) \mid \text{vec}(\mathcal{SR}_2 \circ T_2^\theta) \mid ...\right]$ denote the matrix representing the disparity-compensated super-ray.

Note that in practice, the disparity compensated super-pixels forming a super-ray may not have the same shape and size, thus preventing us from defining $\mathcal{SR} \circ T^\theta$. Therefore, we present in Section V-E a procedure called *super-ray extension* to enforce shape and size consistency of the super-pixels. Hence, for the mathematical derivations, we can assume that this constraint is always satisfied and that the matrix $\mathcal{SR} \circ T^\theta$ is always defined.

The goal is then to reduce the dimensionality of each super-ray by searching for a matrix $\mathbf{M}$ of a lower rank r that will best approximate the input super-ray. The approximation error will be minimized for a given rank if the super-pixels forming the super-ray are well aligned. The problem therefore consists in finding the vector of disparity parameters θ such that the disparity compensated super-ray $\mathcal{SR} \circ T^\theta$ has the best low rank approximation. Formally, given a target rank r, the problem to solve is:

$$\min_{\theta, \mathbf{M}} \left\| \mathcal{SR} \circ T^\theta - \mathbf{M} \right\|_F^2 \text{ s.t. } \text{rank}(\mathbf{M}) = r. \tag{10}$$

This problem is not convex, and there is no theoretical guarantee of convergence. However, in practice (see Fig.2), we have observed that the low rank approximation error ($\left\| \mathcal{SR} \circ T^\theta - \mathbf{M} \right\|_F^2$) of the aligned super-rays (represented by $PSNR_{in}$ in Fig.2) keeps decreasing until reaching a saturation value after a number of iterations given by it_{in} in Fig.2.

B. Proposed algorithm

We solve the problem by alternatively finding $\mathbf{M}$ while fixing θ, and then by updating θ while fixing $\mathbf{M}$. The process is repeated until convergence.

1) Estimating $\mathbf{M}$ for fixed parameters: For fixed parameters θ, the matrix $\mathcal{SR} \circ T^\theta$ is also fixed and the problem in Eq. (10) becomes a simpler low rank approximation problem with a closed form solution. From the singular value decomposition $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ of $\mathcal{SR} \circ T^\theta$, the matrix M of rank r that best approximates $\mathcal{SR} \circ T^\theta$ is:

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}_r\mathbf{V}^\top, \tag{11}$$

where $\mathbf{\Sigma}_r$ contains only the r largest singular values of $\mathbf{\Sigma}$.

2) Estimating θ for fixed $\mathbf{M}$: Searching for the disparity parameters θ that minimize Eq. (10) for a fixed matrix $\mathbf{M}$ is not trivial due to the non linearity of the term $\mathcal{SR} \circ T^\theta$. We instead minimize

$$\min_{\Delta\theta} \left\| \mathcal{SR} \circ T^{\theta+\Delta\theta} - \mathbf{M} \right\|_F^2. \tag{12}$$

Assuming the change $\Delta\theta$ at each iteration to be small, for each super-pixel $\mathcal{SR}_i$ of the super-ray $\mathcal{SR}$, we can approximate this update assuming local linearity as

$$\mathcal{SR}_i \circ T_i^{\theta+\Delta\theta} \approx \mathcal{SR}_i \circ T_i^\theta + \sum_j \frac{\partial[\mathcal{SR}_i \circ T_i^\theta]}{\partial\theta_j}\Delta\theta_j, \tag{13}$$

where the partial derivatives of the transformed super-pixel $\mathcal{SR}_i \circ T_i^\theta$ with respect to each parameter θ_j are given by the chain rule:

$$\frac{\partial[\mathcal{SR}_i \circ T_i^\theta]}{\partial\theta_j} = \frac{\partial\mathcal{SR}_i}{\partial X_i^\theta} \cdot \frac{\partial X_i^\theta}{\partial\theta_j} + \frac{\partial\mathcal{SR}_i}{\partial Y_i^\theta} \cdot \frac{\partial Y_i^\theta}{\partial\theta_j}. \tag{14}$$

In order to rewrite Eq. (13) in matrix form, we note $\mathbf{J_i}(x, y)$ the Jacobian matrix of the transformation T_i^θ with respect to θ, and we define the vector $\mathbf{\Gamma_i}(x, y)$ as follows, for any spatial pixel coordinate (x, y):

$$\mathbf{J_i}(x, y) = \begin{pmatrix} \frac{\partial X_i^\theta}{\partial\theta_1}(x, y) & \dots & \frac{\partial X_i^\theta}{\partial\theta_t}(x, y) \\ \\ \frac{\partial Y_i^\theta}{\partial\theta_1}(x, y) & \dots & \frac{\partial Y_i^\theta}{\partial\theta_t}(x, y) \end{pmatrix}, \tag{15}$$

$$\mathbf{\Gamma_i}(x, y) = \left(\frac{\partial\mathcal{SR}_i}{\partial X_i^\theta}(x, y),\ \frac{\partial\mathcal{SR}_i}{\partial Y_i^\theta}(x, y) \right). \tag{16}$$

In practice, $\mathbf{\Gamma_i}(x, y)$ can be evaluated numerically at each pixel (x, y) by computing the image gradient of $\mathcal{SR}_i$ and by applying the transformation T_i^θ to both the vertical and horizontal components. The Jacobian matrix $\mathbf{J_i}(x, y)$ will depend on the definition of the functions D_x^θ and D_y^θ for a given disparity model.

Let us now define the matrix $\mathbf{G_i}$ as the Jacobian matrix of $\mathcal{SR}_i \circ T_i^\theta$ with respect to θ. Its elements are $[\mathbf{G_i}]_{kj} = \frac{\partial[\mathcal{SR}_i \circ T_i^\theta]}{\partial\theta_j}(x_k, y_k)$, where the row index k ranges over all the pixels in the super-pixel $\mathcal{SR}_i$. From Eqs. (14), (15), and (16), every row of index k of $\mathbf{G_i}$ is computed as:

$$[\mathbf{G_i}]_{k,*} = \mathbf{\Gamma_i}(x_k, y_k)\mathbf{J_i}(x_k, y_k) \tag{17}$$

Now, Eq. (13) can be rewritten as:

$$\mathcal{SR}_i \circ T_i^{\theta+\Delta\theta} \approx \mathcal{SR}_i \circ T^\theta + \mathbf{G_i}\Delta\theta. \tag{18}$$

Given this approximation, the minimisation problem of Eq. (12) is simplified into:

$$\min_{\Delta\theta} \sum_i \left\| \mathbf{G_i}\Delta\theta - \mathbf{R}_i \right\|_F^2, \tag{19}$$

where $\mathbf{R}_i$ is the ith column of the residual matrix $\mathbf{R}$ for the current parameters θ, which is defined by

$$\mathbf{R} = \mathbf{M} - \mathcal{SR} \circ T^\theta. \tag{20}$$

This problem has the following analytical solution:

$$\Delta\theta = \left(\sum_i \mathbf{G_i}^\top \mathbf{G_i} \right)^{-1} \sum_i \mathbf{G_i}^\top \mathbf{R}_i. \tag{21}$$

C. Warping error minimization

Solving Eq. (10) allows to find the optimal parameters θ_{in} to align the super-ray (*forward warping*) so as to minimize the low rank approximation error as

$$\theta_{in} = \text{argmin}_\theta \left\| \mathcal{SR} \circ T^\theta - \mathbf{M} \right\|_F^2 . \tag{22}$$

However it is necessary to compensate this alignment (*i.e. inverse warping*) to recover the original super-ray, which may introduce further errors due to interpolations. For instance, in Fig. 2 we observe that while the low rank approximation error steadily decreases (i.e. *PSNRin* increases) after each iteration of Algorithm 1, the super-ray reconstruction error reaches a minimum value (i.e. maximum of *PSNRout* at iteration it_{out}) and further iterations degrade the result. This reconstruction error increase can be explained by the fact that this error does not only include the low rank approximation error but also includes the inverse warping error which may increase when disparity values increase along the iterations. In order to take

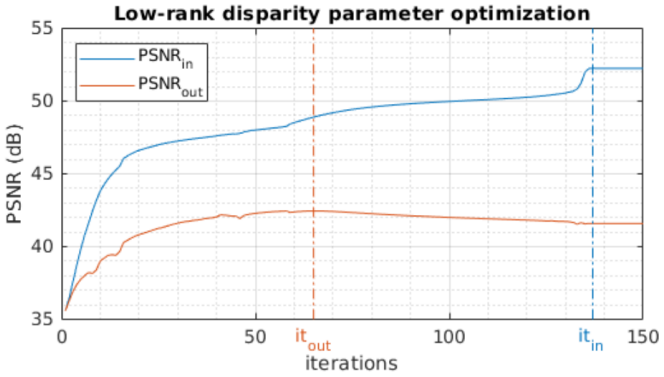

Fig. 2. Low-rank approximation error on the aligned super-ray (PSNRin) and original super-ray reconstruction error (PSNRout) after the three steps, i.e. alignment, low-rank approximation and inverse warping, using Algorithm 1.

into account these errors into our optimisation, we search for the parameters θ_{out} minimizing the reconstruction error after inverse warping, i.e. minimizing

$$\theta_{out} = \text{argmin}_\theta \left\| \mathcal{SR} - \mathbf{M} \circ (T^\theta)^{-1} \right\|_F^2 . \tag{23}$$

We finally retain the parameters θ_{out} minimizing the latter error throughout all iterations.

Indeed, we can see in Fig. 3 that solving Eq. (22) in order to estimate the parameters (θ_{in}) significantly reduces the reconstruction error but that taking into account the inverse warping by minimizing Eq. (23) instead (θ_{out}) reduces it even further.

Accordingly, the performance of our compression scheme significantly increases when taking into account the inverse warping (θ_{out}) compared to discarding it (θ_{in}), as observed on Fig. 4.

The complete algorithm is summarized in Algorithm (1).

D. Algorithm Complexity

Each iteration of Algorithm 1 updates alternatively the matrix $\mathbf{M}$ and the parameters θ. The complexity of the $\mathbf{M}$

Algorithm 1: Low Rank Disparity Estimation (LRDE)

- Initialize θ: $\theta = (0, 0, 0)$
- Initialize θ_{out}: $\theta_{out} = (0, 0, 0)$
- Initialize ϵ: $\epsilon = +\infty$
- For a fixed number of iterations
  - With θ fixed, the optimal matrix $\mathbf{M}$ of rank r is obtained as
    $$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}_r\mathbf{V}^\top, \tag{24}$$
    where $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ is the singular value decomposition (SVD) of $\mathcal{SR} \circ T^\theta$ and $\mathbf{\Sigma}_r$ contains only the r largest singular values of $\mathbf{\Sigma}$.
  - With $\mathbf{M}$ fixed, update θ as $\theta \leftarrow \theta + \Delta\theta$ where $\Delta\theta = \left(\sum_i \mathbf{G_i}^\top \mathbf{G_i}\right)^{-1} \sum_i \mathbf{G_i}^\top \mathbf{R}_i$.
  - Evaluate the reconstruction error
    $$\epsilon_\theta = \left\| \mathcal{SR} - \mathbf{M} \circ (T^\theta)^{-1} \right\|_F^2 \tag{25}$$
  - If the current reconstruction error ϵ_θ is smaller than the minimum error ϵ, update ϵ and θ_{out}
    $$\epsilon \leftarrow \epsilon_\theta, \ \theta_{out} \leftarrow \theta \tag{26}$$
- Output θ_{out}

matrix update is dominated by the singular value decomposition. For a matrix of size $m \times n$, the SVD is computed in $O(\min(m^2n, mn^2))$. In our case, m is the number of pixels and n is the number of views, and in practice, we have $m > n$. Thus, the complexity of this step is $O(mn^2)$.

For updating the parameters $\theta \in \mathbb{R}^{t\times 1}$, we must first compute the matrix $\mathbf{R} \in \mathbb{R}^{m\times n}$ and the matrices $\mathbf{G_i} \in \mathbb{R}^{m\times t}$ for each view i. These computations require a fixed number of operations per pixel and per view, hence giving a complexity of $O(mn)$ since we can ignore the number of parameters t of the model which is a small constant of the problem. Then, the complexity of solving the problem in Eq. (21) is dominated by the computation of $\sum_i \mathbf{G_i}^\top \mathbf{G_i}$ that is performed in $O(mn)$. Note that it results in a $t \times t$ matrix, so the computation of its inverse in Eq. (21) is neglected.

Therefore, the complexity of the parameter update step is $O(mn)$, and the overall complexity of each iteration remains dominated by the SVD step in $O(mn^2)$.

E. Super-ray extension

Note that even after disparity compensation, the aligned super-pixels forming a super-ray may have different shapes and sizes as can be seen in Fig. 5, in particular due to occlusions, which would make the low rank approximation impractical.

To circumvent this problem, each super-ray is extended, i.e. each super-pixel of a super-ray is padded with neighbouring pixels until it matches the area covered by the union of all super-pixels. It might still happen that neighboring pixels do not exist at certain locations, in particular for super-rays located at the border of the light field. While the value of these pixels could be copied from the closest existing neighbor, it

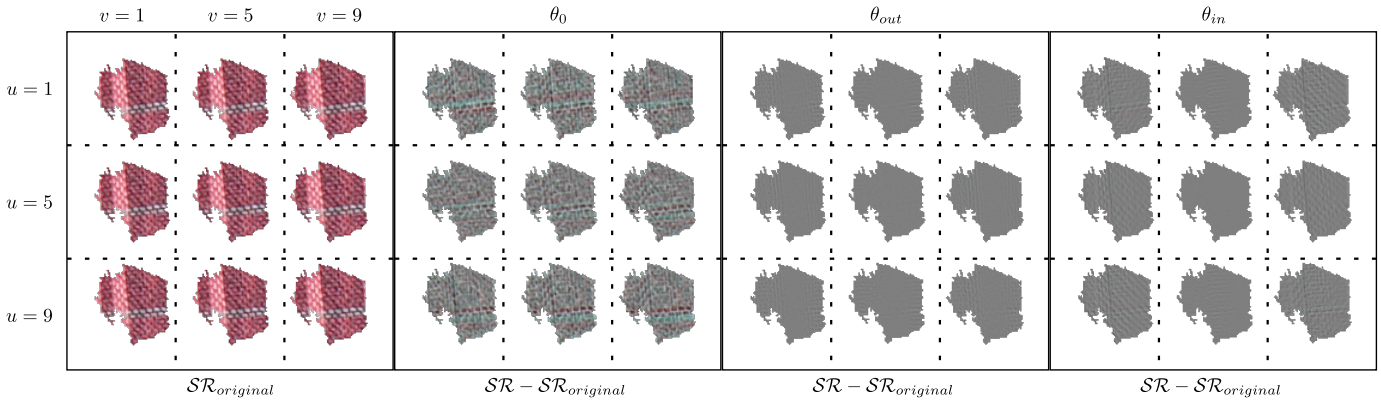

Fig. 3. On the left, we show different views of a super-ray $\mathcal{SR}_{original}$. This super-ray is then reconstructed by performing forward warping, low-rank approximation and inverse warping. On the right, we compare the reconstruction error using different parameters for the disparity compensation: θ_0 (zero disparity), θ_{out} (minimizing Eq. (23)) and θ_{in} (minimizing Eq. (22)).

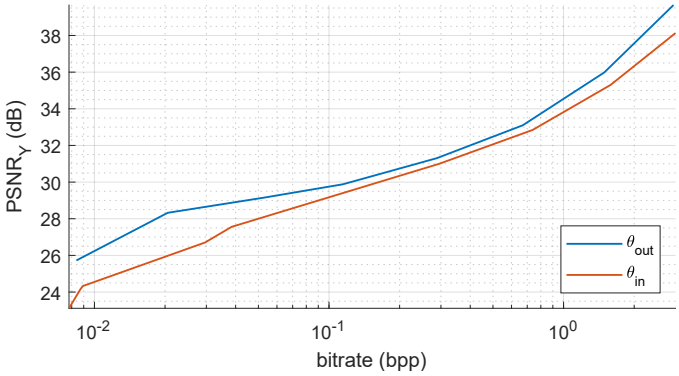

Fig. 4. Performance of our compression scheme with (θ_{out}) and without (θ_{in}) taking into account the inverse warping to evaluate the reconstruction error in Algorithm 1 on the light field Bench [38].

would in general degrade the performance of the ensuing low rank approximation. Instead, the missing pixels are inpainted using a low rank matrix completion so as to yield the lowest error after low-rank approximation.

As all super-pixels now have similar shape and size after extension, the super-ray may be reshaped into a matrix $\mathcal{SR} = [\text{vec}(\mathcal{SR}_1) \mid \text{vec}(\mathcal{SR}_2) \mid ...]$ formed by vectorizing all super-pixels $\mathcal{SR}_i$ for each view i. The low rank matrix completion problem is then posed as the search for the minimum nuclear norm matrix $\widetilde{\mathcal{SR}}$ with entries equal to those of the matrix $\mathcal{SR}$ for the known elements of $\mathcal{SR}$. The problem is mathematically formulated as

$$\min_{\widetilde{\mathcal{SR}}} ||\widetilde{\mathcal{SR}}||_* \text{ s.t. } \forall (i,j) \in \Omega, \widetilde{\mathcal{SR}}_{ij} = \mathcal{SR}_{ij}, \qquad (27)$$

where Ω is the set of indices of the known elements in $\mathcal{SR}$, and $||.||_*$ is the nuclear norm (convex approximation of the rank). This minimization is solved using the Inexact ALM (IALM) technique [39].

VI. Disparity models studied

A. Model of constant disparity per super-ray

We first consider the model that was introduced in [9] where we assume the disparity to be constant within each super-ray which amounts to considering the scene is only composed of planar objects parallel to the camera plane. However, to add more flexibility to the model and to cope with possible inaccuracies in the input light fields, we allow the horizontal and vertical disparities of a super-ray to be different. The parameters to determine are thus $\theta = (d_x, d_y)$, and the constant model is simply defined by $D_x^\theta(x,y) = d_x$ and $D_x^\theta(x,y) = d_y$. The Jacobian matrix $\mathbf{J_i}(x,y)$ of the warping operator T_i^θ defined in Eqs. (6),(7),(8) is then:

$$\mathbf{J}_i(x,y) = \mathbf{J}_i = \begin{pmatrix} u_i - u_c & 0 \\ 0 & v_i - v_c \end{pmatrix}. \qquad (28)$$

Knowing $\mathbf{J_i}$, the constant disparity values d_x and d_y can be solved with Algorithm (1), using the definition of $\mathbf{G_i}$ in Eq. (17).

For all light fields, d_x and d_y are both initialized to 0 in each super-ray.

B. Affine disparity model per super-ray

Now, we consider a finer model where the disparity can vary within each super-ray according to an affine function:

$$D_x^\theta(x,y) = D_y^\theta(x,y) = \alpha \cdot x + \beta \cdot y + \gamma, \qquad (29)$$

parameterized by $\theta = (\alpha, \beta, \gamma)$ which now allows the planar objects in the scene to be inclined with respect to the camera plane. We assume the horizontal and vertical disparities are equal to reduce the number of parameters in θ from 6 to 3.

The Jacobian matrix $\mathbf{J_i}(x,y)$ for the corresponding warping operator T_i^θ is then:

$$\mathbf{J_i}(x,y) = \begin{pmatrix} x \cdot (u_i - u_c) & y \cdot (u_i - u_c) & (u_i - u_c) \\ x \cdot (v_i - v_c) & y \cdot (v_i - v_c) & (v_i - v_c) \end{pmatrix}. \qquad (30)$$

Similarly to the previous case where the disparity is assumed to be constant, the algorithm iteratively proceeds as described in Algorithm 1.

For all light fields, θ is initialized to $(0,0,0)$ in each super-ray, corresponding to $D_x^\theta(x,y) = D_y^\theta(x,y) = 0$.

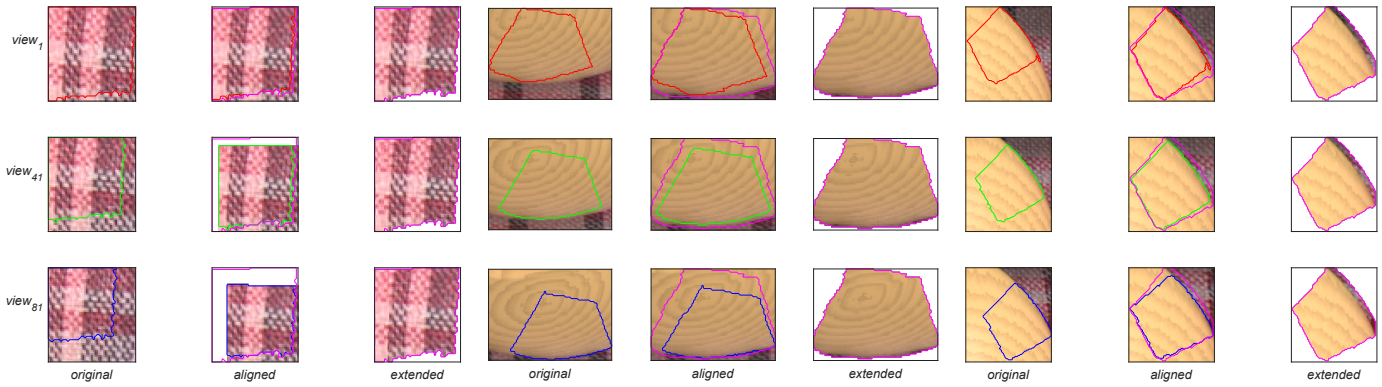

Fig. 5. Example of super-ray alignment and extension for the light field StillLife [40]. The red, green and blue outline respectively represent the super-pixel boundaries in views 1, 41 and 81. The magenta outline represents the union of the areas covered by each super-pixel which is used to delineate the pixels to include in the extended super-pixels.

VII. Low Rank Approximation

In order to reduce the amount of data to be transmitted, we first reduce the dimensionality of the input light field data by using a low rank approximation. The goal is to compact the energy of the light field data in as few components as possible, a.k.a eigen images as we will see in the sequel that the low rank approximation method relies on a singular value decomposition. All the matrices corresponding to the different aligned super-rays

$$\mathcal{SR}_{k,aligned} = \mathcal{SR}_k \circ T^{\theta_k}, \tag{31}$$

are stacked in a matrix $\mathbf{X}$ of dimension $\mathbb{R}^{m\times n}$, where n is the number of views and m is the number of pixels per view of the concatenated aligned super-rays. In other words, the matrix $\mathbf{X}$ contains all the pixels of all the views that are locally aligned super-ray per super-ray. The matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ is factorized into the product of a matrix $\mathbf{B} \in \mathbb{R}^{m\times r}$ and a coefficient matrix $\mathbf{C} \in \mathbb{R}^{r\times n}$, with $r \leq n$, as

$$\underset{\mathbf{B},\mathbf{C}}{\arg\min} \|\mathbf{X} - \mathbf{BC}\|_F^2, \tag{32}$$

where $\|.\|_F$ is the Frobenius norm. Optimal factorization is obtained from the singular value decomposition (SVD) of $\mathbf{X}$ into $\mathbf{U\Sigma V}^\top$. Assuming the singular values in $\mathbf{\Sigma}$ are in decreasing order, we take $\mathbf{B}$ as the r first columns of $\mathbf{U\Sigma}$, and $\mathbf{C}$ as the r first rows of $\mathbf{V}^\top$.

The set of super-rays $\{\mathcal{SR}\}$ and by extension the entire light field can thus be approximated by a linear combination of columns of $\mathbf{B}$, therefore significantly reducing the amount of data at the cost however of an approximation error.

As each column of $\mathbf{B}$ is a scaled (left-)eigen vector of $\mathbf{X}$, it can be represented as an *eigen image* of the light field. Similarly, the columns of the sub-matrix $\mathbf{B_k}$, obtained by selecting in $\mathbf{B}$ the lines corresponding to $\mathcal{SR}_{k,aligned}$ in $\mathbf{X}$, can be represented as *eigen super-pixels*, which helps understanding the impact of the disparity compensation on the low rank approximation.

The eigen super-pixels corresponding to the first 3 columns of $\mathbf{B_k}$ are represented in Fig. 6 when using different methods to obtain the parameters θ_k for the disparity compensation of $\mathcal{SR}_k$: using no disparity compensation (θ_0), minimizing the low rank approximation error in Eq. (22) (θ_{in}) or minimizing the super-ray reconstruction error in Eq. (23) (θ_{out}).

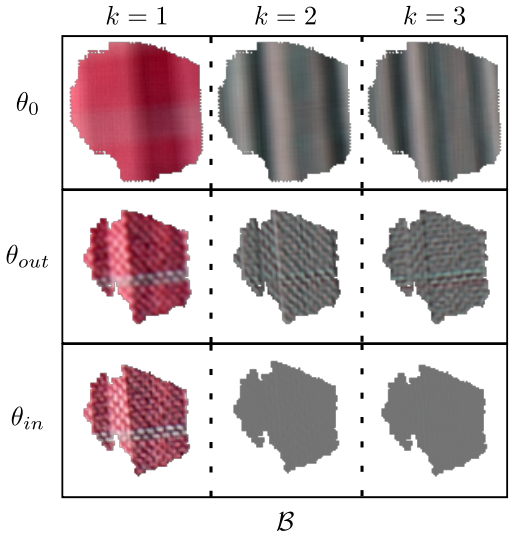

Fig. 6. Visualization of the first, second and third column of B_k as images using different parameters θ for the disparity compensation: θ_0 (zero disparity), θ_{out} (minimizing Eq. (23)) and θ_{in} (minimizing Eq. (22)).

We observe that minimizing Eq. (22) encourages the information in $\mathbf{B_k}$ to be concentrated in the first columns, effectively reducing the low rank approximation error when compared to using no disparity compensation. Minimizing Eq. (23) instead allows a smaller super-ray reconstruction error, as seen in Fig. 3, at the cost of a slightly worse low rank approximation error.

VIII. Compression Scheme

The above low rank approximation computes two matrices: a matrix $\mathbf{B}$ containing column-wise concatenated extended super-pixels and a matrix $\mathbf{C}$ containing the weighting coefficients. In order to encode the matrix $\mathbf{B}$ using HEVC, the data in each column l is re-arranged into an image $\mathbf{B}^l$ called *eigen-image* by stitching neighboring pixels. However due to

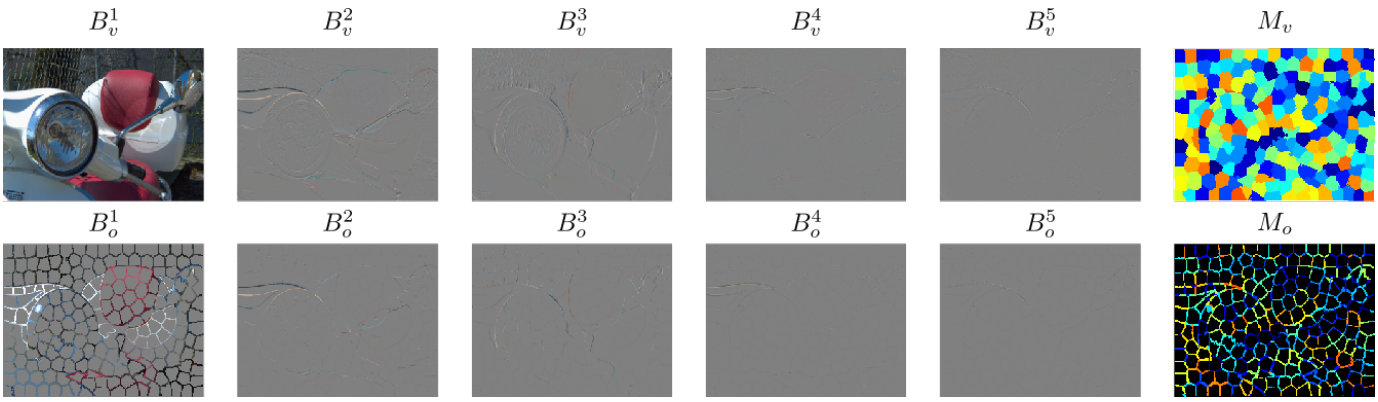

Fig. 7. First 5 eigen-images and segmentation maps for the visible ($\mathbf{B}_v^l$, $\mathbf{M}_v$) and occluded ($\mathbf{B}_o^l$, $\mathbf{M}_o$) sets of pixels ("Vespa" [41]).

the alignment and extension processes, neighboring super-pixels may overlap where occlusions and disocclusions occur, rendering the stitching impractical.

To overcome this difficulty, instead we create two images $\mathbf{B}_v^l$ and $\mathbf{B}_o^l$, by respectively grouping pixels that are visible and occluded in the central view. Non-overlapping pixels are assigned to $\mathbf{B}_v^l$ as they are visible in the central view without ambiguity, while overlapping pixels are assigned to either $\mathbf{B}_v^l$ or $\mathbf{B}_o^l$ depending on the disparity associated to each pixel. For two overlapping pixels $p_i(x,y)$ and $p_j(x,y)$ from super-pixels $\mathcal{SR}_i$ and $\mathcal{SR}_j$, with respective disparity values $d_i = \alpha_i \cdot x + \beta_i \cdot y + \gamma_i$ and $d_j = \alpha_j \cdot x + \beta_j \cdot y + \gamma_j$, if $d_i > d_j$, then p_i is assumed to belong to the foreground of the scene. Thus, p_i is visible in the central view and assigned to $\mathbf{B}_v^l$ while p_j is assigned to $\mathbf{B}_o^l$ instead.

Correspondingly, two segmentation maps $\mathbf{M}_v$ and $\mathbf{M}_o$ store the index of the super-ray associated with each pixel in $\mathbf{B}_v^l$ and $\mathbf{B}_o^l$ respectively. These maps are used to recover the visible and occluded parts of each super-pixel. Examples of eigen images $\mathbf{B}_v^l$ and $\mathbf{B}_o^l$ with their associated segmentation maps $\mathbf{M}_v$ and $\mathbf{M}_o$ are illustrated in Fig.7.

To efficiently encode the $\{\mathbf{B}_v\}$ and $\{\mathbf{B}_o\}$ sets we take full advantage of the existing correlations between corresponding images in $\{\mathbf{B}_v\}$ and $\{\mathbf{B}_o\}$ which are visible in Fig. 7. First, in order to further increase these correlations, the missing information in $\{\mathbf{B}_o\}$ images (black regions of $\mathbf{M}_o$ in Fig. 7) is replaced by copying the collocated regions of the corresponding $\{\mathbf{B}_v\}$ images. Then the images in $\{\mathbf{B}_v\}$ and $\{\mathbf{B}_o\}$ are quantized on 16 bits, interleaved as $\mathbf{B}_v^1$, $\mathbf{B}_o^1$, ..., $\mathbf{B}_v^r$, $\mathbf{B}_o^r$ (where r is the approximation rank) and encoded with HEVC using the "IP" (I-frame followed by P-frame) Group of Picture (GOP) structure to fully exploit the redundancy between collocated pixels in consecutive $\mathbf{B}_v^k$ (visible) and $\mathbf{B}_o^k$ (occluded) frames.

The entries of the matrix $\mathbf{C}$ are only quantized on 16 bits using fixed length encoding and transmitted as such since the corresponding cost is quite negligible.

It is also necessary to transmit as side information the three disparity parameters per super-ray as well as the two segmentation maps $\mathbf{M}_v$ and $\mathbf{M}_o$. The red, green and blue channels of a three-channel image are respectively filled with $\mathbf{M}_v$, $\mathbf{M}_o$ and $\mathbf{M}_o$ and this image is compressed losslessly using the FLIF coder [8]. The disparity information (three parameters α, β and γ per super-ray) is encoded using a fixed length code on 32 bits, which is quite negligible given the small number of values to be transmitted (in the experiments we considered 200 super-rays for natural light fields and 240 for the synthetic ones).

IX. Light field reconstruction

In order to reconstruct the entire light field, the sequence of decoded eigen images is first deinterleaved to recompose the $\{\tilde{\mathbf{B}}_v\}$ and $\{\tilde{\mathbf{B}}_o\}$ sets which are then merged to form the decoded low rank matrix $\tilde{\mathbf{B}}$ containing the eigen images of all the aligned super-rays. This matrix $\tilde{\mathbf{B}}$ is multiplied by the matrix $\tilde{\mathbf{C}}$ to recover the matrix $\tilde{\mathbf{X}}$ composed of the aligned super-rays stacked vertically. The aligned super-rays are then extracted from $\tilde{\mathbf{X}}$ using the segmentation maps $\mathbf{M}_v$ and $\mathbf{M}_o$ to locate and merge the visible and occluded parts of each super-ray. The unaligned super-rays are recovered by performing inverse disparity compensation $T^{\theta_k^{-1}}$ on each aligned super-ray $\tilde{\mathcal{SR}}_{k,aligned}$ using its associated disparity model parameters θ_k.

$$\tilde{\mathcal{SR}}_k = \tilde{\mathcal{SR}}_{k,aligned} \circ T^{\theta_k^{-1}}. \tag{33}$$

Finally, the light field is progressively reconstructed by mapping each view of a super-ray to the corresponding view of the light field, one super-ray at a time. In case of overlap between pixels reprojected from different super-rays, the pixel of higher disparity is kept.

Depending on the complexity of the scene, on the number of objects and depth layers, regions in the light field may be occluded by several objects. The corresponding pixels are referred to as *multiply-occluded* pixels. Two sets of eigen images $\{\mathbf{B}_o^l\}_l$ may not be sufficient to represent all *multiply-occluded* pixels. One alternative would be to transmit additional sets $\{\mathbf{B}_o^l\}_l$ to represent these pixels. However, the number of such pixels is quite limited and the corresponding matrix would be very sparse. Instead, we inpaint the corresponding pixels in the reconstructed light field using the same low rank matrix completion method used for super-ray extension presented in Eq. (27) of Section V-E.

X. Experimental Results

The performances of the proposed disparity estimation and light field compression methods have been evaluated for the luminance component of light fields shown in Fig.9 and coming from the HCI [40], INRIA [38] and ICME 2016 Grand Challenge [41] datasets. The HCI dataset contains synthetic light fields with 9×9 views of 768×768 pixels and the INRIA and ICME datasets contain light fields captured by a Lytro Illum camera from which we use the 9×9 central sub-aperture images cropped to 616×424 pixels to remove extremely noisy or black pixels. The Lytro light fields have been decoded using the Matlab Light Field Toolbox v0.4 [42] with gamma correction.

The HLRA and proposed LLRA methods have been evaluated for varying rank values $r = \{1, 3, 5, 10, 15, 30, 60\}$ and HEVC quality parameters $QP \in \{5, 10, 15, 20, 30, 40, 50\}$ and we retain the (r,QP) pairs corresponding to the points on the convex envelop of the rate-distortion plots. For the HEVC Lozenge method [11], the QP parameter has been set to $\{10, 14, 17, 20, 23, 26\}$. The disparity maps used for LLRA and JPEG Pleno have been generated using [10]. For JPEG Pleno, we used the publicly available configuration file of the 'greek' LF as it is designed for 9×9 LFs. To closely match the JPEG Pleno Test Conditions [43], the Bjontegaard metrics have been evaluated for bitrates ranging from 0.005 bpp to 0.75 bpp.

A. Performance analysis of the disparity compensation models

We first compare in Fig.8 and Tab. I the performance of the full compression scheme using the constant model as proposed in [9] and the proposed affine disparity models.

At lower bitrates (≤ 0.1 bpp), corresponding to small rank values $(1, 3, 5)$ where the information sent to the decoder is limited, we observe increased performance when using an affine model instead of a constant model to perform the disparity compensation. This is explained by the fact that the affine model better aligns the super-pixels of each super-ray, which reduces the low-rank approximation error. At higher bitrates (> 0.1 bpp), corresponding to higher rank values $(10, 15, 30, 60)$ the lesser alignment of super-pixels is compensated by the additional information allowed to be sent.

In order to show the interest of using the low rank prior in the disparity estimation, we have also tested an alternative approach where the affine model is directly learnt to fit an input disparity map (see 'affine, input disparity' in Fig. 8 and Tab. I). Given the disparity map $D(x, y, u, v)$, for each pixel (x, y, u, v) within a super-ray $\mathcal{SR}$, the parameters $\theta = (\alpha, \beta, \gamma)$ are obtained by solving the following linear regression:

$$\min_{\alpha,\beta,\gamma} \left\| \sum_{(x,y,u,v)\in\mathcal{SR}} (D(x, y, u, v) - (\alpha \cdot x + \beta \cdot y + \gamma)) \right\|_F^2 . \quad (34)$$

For natural light fields (Fig. 8, bottom), both affine models perform similarly at lower bitrates ($\leq$ 0.1 bpp) while we observe a significant gain in performance using a low-rank prior at higher bitrates (> 0.1 bpp). As discussed in Section V-C, Eq. (10) drives the disparity parameter estimation by minimizing the low rank approximation of a super-ray after warping, which tends to align super-pixels, but the inverse warping performed to recover the super-ray is not accounted for and introduces interpolation errors. As the rank of the low-rank approximation increases, these interpolation errors become more penalizing and it becomes more valuable to limit the alignment by minimizing the super-ray reconstruction error as expressed in Eq. (23). It is worth mentioning that the same behaviour is observed using the constant model with the low-rank prior parameter estimation method which explains the similar performance using the constant or the affine model for high bitrates.

For synthetic light fields (Fig. 8, top) the affine model shows systematic improvement over the constant model. However, depending on the light field, better results may be obtained by either fitting an input disparity map (following (34)) or using the low rank prior procedure (Alg.1) to estimate the parameters of the affine model. The better results obtained by fitting the model on an input disparity map observed for some synthetic data can be explained by the fact that for synthetic data we used ground truth disparity. In the case of natural light fields (Fig. 8, bottom), the disparity is estimated using [10], and in this case the model parameter estimation with the low rank prior gives better results at high bit rates.

Hence, when comparing our compression scheme to other state-of-the-art methods we use an affine disparity model with disparity parameters estimated using the low rank prior.

Reference Method	Constant low rank prior	Affine input disparity
Bench	-15.96 %	-24.78 %
Fruits	-6.4 %	2.07 %
Toys	-26.63 %	-8.33 %
Fountain & Vincent 2	-27.79 %	-1.36 %
Friends 1	-59.76 %	-11.10 %
Stone Pillars Inside	-20.74 %	-18.04 %
Vespa	-31.82 %	-0.18 %
Buddha	-25.40 %	7.04 %
Butterfly	-57.47 %	-27.24 %
StillLife	-37.61 %	19.31 %

TABLE I
BJONTEGAARD RATE SAVINGS FOR LLRA USING AN AFFINE MODEL WITH LOW-RANK PRIOR PARAMETER ESTIMATION AGAINST USING EITHER A CONSTANT MODEL WITH LOW RANK PRIOR [9] OR AN AFFINE MODEL FITTED FROM AN INPUT DISPARITY MAP.

B. Performance analysis of the complete scheme

We compare the performance of our method (LLRA), HLRA [5] and HEVC-Lozenge [11] against the JPEG Pleno anchor [12] in the 4D prediction mode (i.e. WASP method). Using the Bjontegaard-rate metric reported in Tab. II we observe that HLRA outperforms the other methods on natural light fields, however we obtain slightly better performance for LLRA on synthetic light fields (Buddha, Butterfly). Indeed we observe in Fig.10 that our method compares favorably with the other presented methods, HLRA included, but that a significant gain can be observed at medium bitrates for

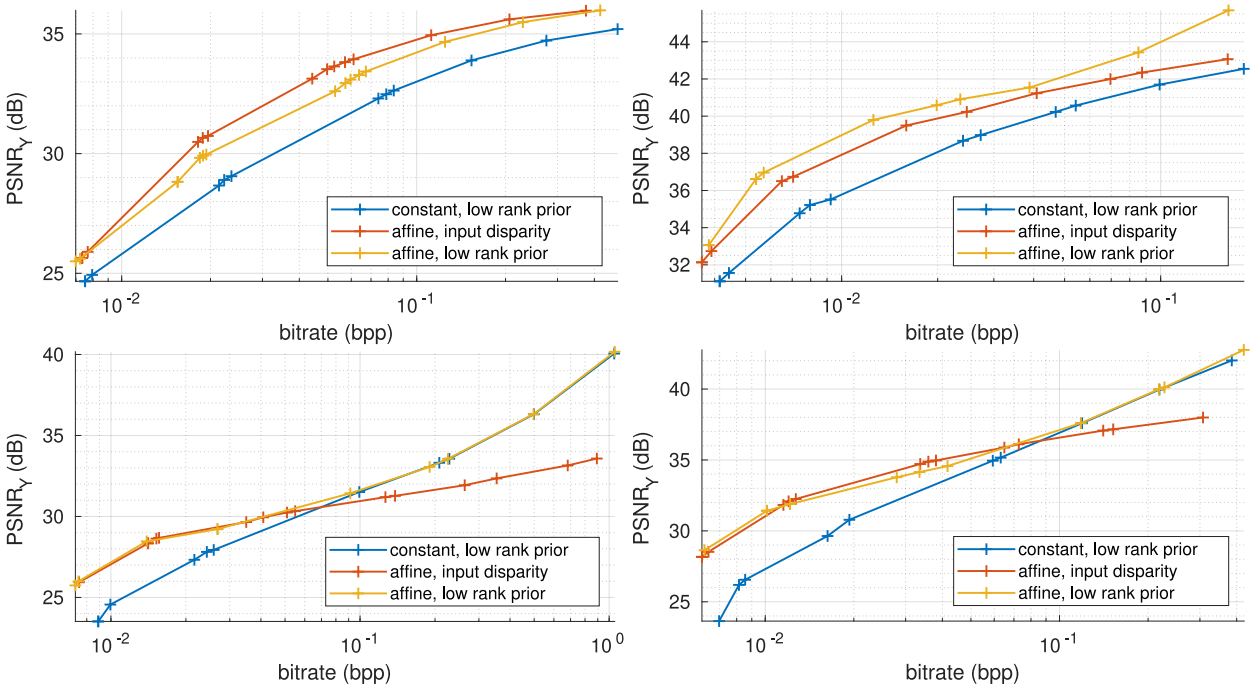

Fig. 8. PSNR-Rate performance of the different disparity compensation models and parameter estimation methods with two synthetic light fields on the top row (StillLife (left), Butterfly (right) [40]) and two natural light fields on the bottom row (Bench (left), Fountain_Vincent_2 (right) [41]).

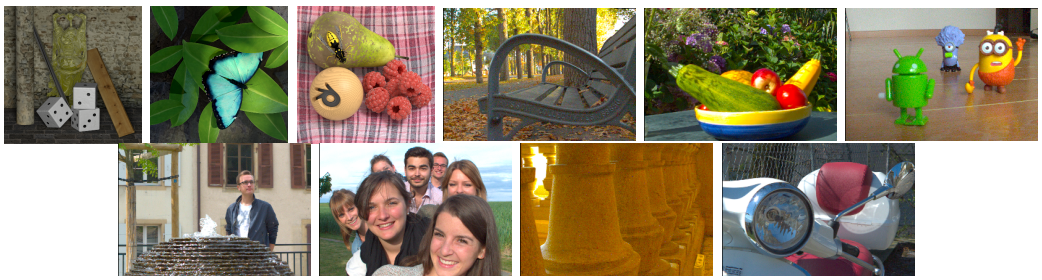

Fig. 9. Test light fields. Top row (first three): HCI (Buddha, Butterfly, StillLife) [40]. Top row (last three): INRIA Dataset (Bench, Fruits, Toys) [38]; Bottom row: ICME Dataset (Fountain_Vincent_2, Friends_1, StonePillars, Vespa) [41].

synthetic light fields. This difference in performance can be explained by the fact that natural light fields usually exhibit small baselines. Thus we see smaller PSNR improvements by performing a local alignment (LLRA) compared to a global alignment (HLRA). It is also necessary to send additional data to the decoder ($\mathbf{B}_o^l$, $\mathbf{M}_v$, $\mathbf{M}_o$ and θ) which penalizes LLRA. However synthetic light fields have a larger baseline and the PSNR gains are significant enough to compensate for the additional data. We can also observe that the proposed scheme outperforms the JPEG pleno anchor and HEVC-Lozenge for natural light fields. For synthetic light fields with larger baselines, the proposed scheme gives the best performances for small and medium bit rates, while it can be outperformed by JPEG-Pleno at high bit rates.

XI. Conclusion

In this paper we have presented a compression scheme for light fields using super-ray based local low rank models. A model to represent disparity within a super-ray as an affine function of spatial coordinates was presented and an algorithm was defined to estimate the optimal parameters to perform super-ray disparity compensation so as to yield the lowest approximation error for a given rank. The experimental results show that the disparity parameters optimized with the low rank prior significantly improves the coding performances compared to directly fitting the affine model's parameters to the input disparity map. Furthermore, using an affine disparity model instead of a constant disparity value substantially improves our results in spite of the increased number of parameters. Finally, despite the need for additional side information, our method compares favorably with the other state-of-the-art methods assessed and can outperform them especially for low and medium bit rates.

References

[1] C. Conti, P. Nunes, and L. D. Soares, "HEVC-based light field image coding with bi-predicted self-similarity compensation," in *IEEE Int.*

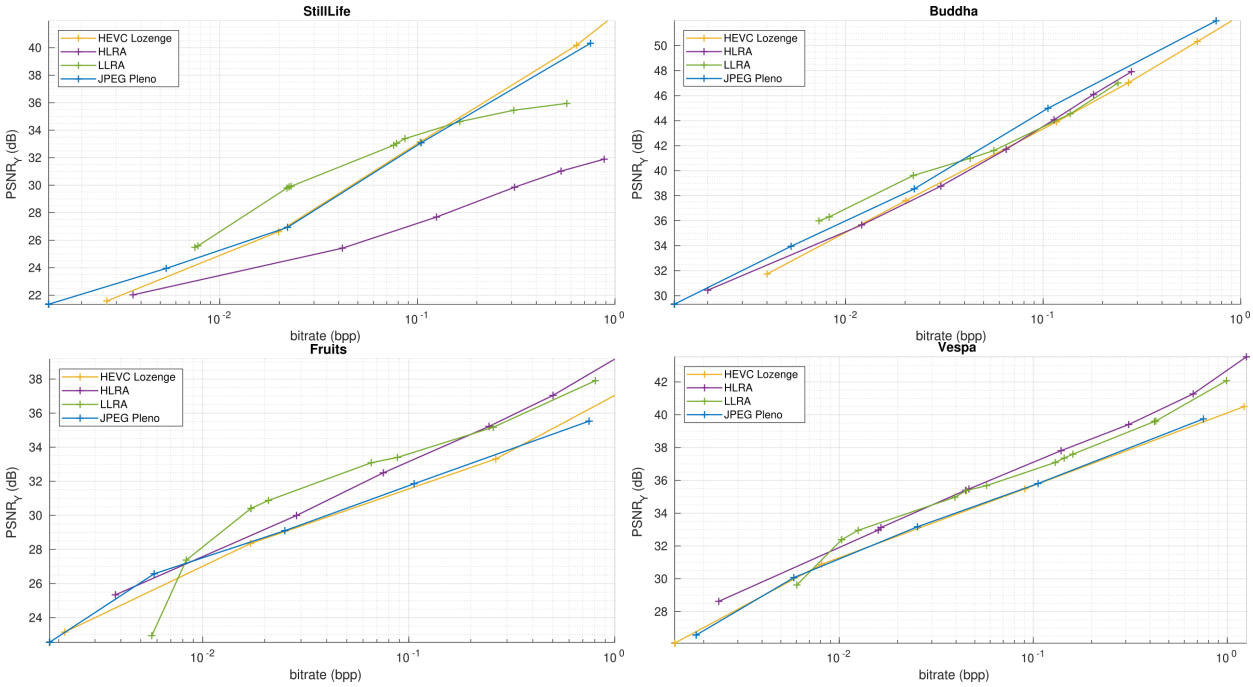

Fig. 10. PSNR-Rate performance of all methods for two synthetic light fields on the top row (StillLife (left), Buddha (right) [40]) and two natural light fields on the bottom row (Fruits (left) [38], Vespa (right) [41]).

	HEVC Lozenge	HLRA	LLRA
Bench	-4.05 %	**-65.27** %	-48.95 %
Fruits	9.99 %	-34.47 %	**-36.72** %
Toys	11.99 %	-71.91 %	**-92.61** %
Fountain Vincent 2	4.22 %	**-46.47** %	-41.31 %
Friends 1	2.98 %	**-62.61** %	-49.49 %
Stone Pillars Inside	-35.82 %	**-87.53** %	-75.54 %
Vespa	0.02 %	**-40.67** %	-35.74 %
Buddha	33.90 %	28.22 %	6.83 %
Butterfly	-33.47 %	-26.58 %	**-37.72** %
StillLife	6.78 %	335.51 %	**-34.23** %
Bench	-0.13 dB	**1.63** dB	1.36 dB
Fruits	-0.17 dB	1.10 dB	**1.49** dB
Toys	0.57 dB	1.25 dB	**1.42** dB
Fountain Vincent 2	-0.08 dB	**1.60** dB	1.32 dB
Friends 1	0.01 dB	**2.40** dB	1.71 dB
Stone Pillars Inside	-0.22 dB	**1.18** dB	0.90 dB
Vespa	-0.00 dB	**1.20** dB	0.98 dB
Buddha	-1.07 dB	-0.89 dB	-0.14 dB
Butterfly	1.30 dB	1.09 dB	**1.48** dB
StillLife	-0.07 dB	-4.56 dB	**0.74** dB

TABLE II
BJONTEGAARD RATE SAVINGS AND PSNR GAINS FOR THE LOZENGE, HLRA AND LLRA COMPRESSION SCHEMES WITH RESPECT TO THE JPEG PLENO VM 2.0 ANCHOR.

Conf. Multimed. Expo Workshops (ICMEW), Jul. 2016.

[2] R. Monteiro, L. Lucas, C. Conti, P. Nunes, N. Rodrigues, S. Faria, C. Pagliari, E. da Silva, and L. Soares, "Light field hevc-based image coding using locally linear embedding and self-similarity compensated prediction," in *IEEE Int. Conf. Multimed. Expo Workshops (ICMEW)*. IEEE, 2016, pp. 1–4.

[3] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Efficient intra prediction scheme for light field image compression," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 539–543.

[4] W. Ahmad, M. Ghafoor, S. A. Tariq, A. Hassan, M. Sjostrom, and R. Olsson, "Computationally efficient light field image compression using a multiview hevc framework," *IEEE access*, vol. 7, pp. ss. 143002–143014, 2019.

[5] X. Jiang, M. Le Pendu, R. Farrugia, and C. Guillemot, "Light field compression with homography-based low-rank approximation," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 1132–1145, Oct. 2017.

[6] M. Hog, N. Sabater, and C. Guillemot, "Super-rays for efficient light field processing," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 1187—-199, Oct. 2017.

[7] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. IEEE Int. Conf. on Computer Vision, ICCV*, 2003, pp. 10–17.

[8] Jon Sneyers and Pieter Wuille, "FLIF: Free lossless image format based on MANIAC compression," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016.

[9] E. Dib, M. Le Pendu, X. Xiang, and C. Guillemot, "Super-ray based low rank approximation for light field compression," in *Data Compression Conference (DCC)*, March 2019.

[10] X. Jiang, J. Shi, and C. Guillemot, "A learning based depth estimation framework for 4d densely and sparsely sampled light fields," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 2257–2261.

[11] M. Rizkallah, T. Maugey, C. Yaacoub, and C. Guillemot, "Impact of light field compression on focus stack and extended focus images," in *European Signal Processing Conf. (EUSIPCO)*, Aug. 2016, pp. 898–902.

[12] ISO/IEC JTC 1/SC29/WG1 JPEG, "Verification model software version 2.0 on jpeg pleno light field coding," Doc. N82046, 2018.

[13] Y. Li, M. Sjostrom, and R. Olsson, "Coding of focused plenoptic contents by displacement intra prediction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 7, pp. 1308–1319, Jui. 2016.

[14] R. J. S. Monteiro, P. J. L. Nunes, N. M. M. Rodrigues, and S. M. M. Faria, "Light field image coding using high-order intra block prediction," *J. on Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1120–1131, Oct. 2017.

[15] R. Zhong, S. Wang, B. Cornelis, Y. Zheng, J. Yuan, and A. Munteanu, "Efficient directional and l1-optimized intra-prediction for light field image compression," in *IEEE Int. Conf. Image Process. (ICIP)*, Sept. 2017, pp. 1172–1176.

[16] Y.-H. Chao, G. Cheun, and A. Ortega, "Pre-demosaic light field image compression using graph lifting transform," in *IEEE Int. Conf. Image Process. (ICIP)*, Sept. 2017, pp. 3240–3244.

[17] C. Conti, L. D. Soares, and P. Nunes, "Light field coding with field-of-view scalability and exemplar-based interlayer prediction," *IEEE Transactions on Multimedia*, vol. 20, no. 11, pp. 2905–2920, Nov 2018.

[18] C. Perra and P. Assuncao, "High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement," in *IEEE Int. Conf. Multimed. Expo Workshops (ICMEW)*, Jul. 2016.

[19] D. Liu, P. An, R. Ma, W. Zhan, X. Huang, and A. A. Yahya, "Content-based light field image compression method with gaussian process regression," *IEEE Transactions on Multimedia*, 2019.

[20] C. Jia, Y. Yang, X. Zhangy, X. Zhang, S. Wangx, S. Wang, and S. Ma, "Optimized inter-view prediction based light field image compression with adaptive reconstruction," in *IEEE Int. Conf. on Image Processing, ICIP*, 2017.

[21] W. Ahmad, R. Olsson, and M. Sjostrom, "Interpreting plenoptic images as multiview sequences for improved compression," in *IEEE Int. Conf. Image Process. (ICIP)*, 2017.

[22] Dong Liu, Lizhi Wang, Li Li, Zhiwei Xiong, Feng Wu, and Wenjun Zeng, "Pseudo-sequence-based light field image compression," in *IEEE Int. Conf. Multimed. Expo Workshops (ICMEW)*. IEEE, 2016, pp. 1–4.

[23] S. Kundu, "Light field compression using homography and 2d warping," in *IEEE Int. Conf. Acoust., Speech Signal Process (ICASSP)*, Mar. 2012, pp. 1349–1352.

[24] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 34, no. 6, pp. 381–395, Jun. 1981.

[25] Y. Li, M. Sjostrom, R. Olsson, and U. Jennehag, "Scalable coding of plenoptic images by using a sparse set and disparities," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 80–91, Jan. 2016.

[26] L. Li, Z. Li, B. Li, D. Liu, and H. Li, "Pseudo-sequence-based 2-d hierarchical coding structure for light-field image compression," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 1107–1119, Oct. 2017.

[27] S. Zhao and Z. Chen, "Light field image coding via linear approximation prior," in *IEEE Int. Conf. on Image Processing, ICIP*, 2017.

[28] X. Jiang, M. Le Pendu, and C. Guillemot, "Light fields compression using depth image based view synthesis," in *Hot3D workshop held jointly with IEEE Int. Conf. Multimed. Expo, ICME*, Jul. 2017.

[29] I. Tabus, P. Helin, and P. Astola, "Lossy compression of lenslet images from plenoptic cameras combining sparse predictive coding and jpeg 2000," in *IEEE Int. Conf. Image Process. (ICIP)*. IEEE, 2017, pp. 4567–4571.

[30] T.-H. Tran, Y. Baroud, Z. Wang, S. Simon, and D. Taubman, "Light-field image compression based on variational disparity estimation and motion-compensated wavelet decomposition," in *IEEE International Conference on Image Processing (ICIP)*, Sept. 2017, pp. 3260–3264.

[31] J. Chen, J. Hou, and L.-P. Chau, "Light field compression with disparity-guided sparse coding based on structural key views," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 314–324, Jan. 2018.

[32] F. Hawary, C. Guillemot, D. Thoreau, and G. Boisson, "Scalable light field compression scheme using sparse reconstruction and restoration," in *IEEE International Conference on Image Processing (ICIP)*, Sept. 2017, pp. 3250–3254.

[33] E. Dib, M. L. Pendu, and C. Guillemot, "Light field compression using fourier disparity layers," in *2019 IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 3751–3755.

[34] W. Ahmad, S. Vagharshakyan, M. Sjöström, A. Gotchev, R. Bregovic, and R. Olsson, "Shearlet transform based prediction scheme for light field compression," in *International Data Compression Conference (DCC)*, Mar. 2018, p. 396.

[35] R. Verhack, T. Sikora, G. Van Wallendael, and P. Lambert, "Steered mixture-of-experts for light field images and video: Representation and coding," *IEEE Transactions on Multimedia*, 2019.

[36] M. B. de Carvalho, M. P. Pereira, G. Alves, E. A. B. da Silva, C. L. Pagliari, F. Pereira, and V. Testoni, "A 4d dct-based lenslet light field codec," in *IEEE International Conference on Image Processing (ICIP)*, Oct. 2018, pp. 435–439.

[37] M. Rizkallah, X. Su, T. Maugey, and C. Guillemot, "Graph-based transforms for predictive light field compression based on super-pixels," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP*, 2018.

[38] "INRIA Lytro image dataset," https://www.irisa.fr/temics/demos/lightField/LowRank2/datasets/datasets.html.

[39] Z. Lin, M. Chen, L. Wu, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," Tech. Rep., University of Illinois at Urbana-Champaign, 2009.

[40] S. Wanner, S. Meister, and B. Goldluecke, "Datasets and benchmarks for densely sampled 4D light fields," in *VMV Workshop*, 2013, pp. 225–226.

[41] "ICME 2016 Grand Challenge dataset," http://mmspg.epfl.ch/EPFL-light-field-image-dataset.

[42] D.G. Dansereau, "Light Field Toolbox for Matlab," 2015.

[43] ISO/IEC JTC 1/SC29/WG1 JPEG, "Jpeg pleno common test conditions," Doc. N81022, 2018.